

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Bayesian Modeling Approaches for Temporal Dynamics in RNA-seq Data

---

Sunghee Oh and Seongho Song

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.73062>

---

## Abstract

Analysis of differential expression has been a central role to address the variety of biological questions in the manner to characterize abnormal patterns of cellular and molecular functions for last decades. To date, identification of differentially expressed genes and isoforms has been more intensively focused on temporal dynamics over a series of time points. Bayesian strategies have been successfully employed to uncover the complexity of biological interest with the methodological and analytical perspectives for the various platforms of high-throughput data, for instance, methods in differential expression analysis and network modules in transcriptome data, peak-callers in ChipSeq data, target prediction in microRNA data and meta-methods between different platforms. In this chapter, we will discuss how our methodological works based on Bayesian models address important questions to arise in the architecture of temporal dynamics in RNA-seq data.

**Keywords:** hierarchical Dirichlet Bayesian mixture model, Poisson Gamma autoregressive model, temporal dynamics, RNA-seq

---

## 1. Introduction

The differential expression analysis across external conditions (e.g. drug treatments, or between- or within-cell/tissue types) in stimuli-response data has long been crucial part on clinical applications [1–11]. The primary goal of these studies is to target therapeutic effects on genes and their pathways that are highly associated with the alterations between different conditions, corresponding underlying biological mechanisms, and condition-specific molecular processes from microarray until recent RNA-seq platform [1–3, 5, 6, 9–43]. Such substantial effects on transcriptome data involved by various types of human diseases have

significantly addressed fundamental issues characterized by biological phenomena in transcriptional regulation. For instance, examination of classification of subtypes on hereditary breast and ovary cancer [44–46], reciprocal phylogenetic conservation and heterogeneity between closest animal models of aging and depression in brain tissues [47–49], identification of differential expression on enzyme effect in Gaucher’s disease across distinct three tissues [14, 50], developmental transient patterns in mouse embryonic stem cells in pre-frontal cortex and limb tissues [13, 51], and et cetera.

The characterization based on Bayesian strategies for whole-genome wide transcriptome data and other types of Seq data has been successfully addressed on the variety of questions to arise in biomedical community [8, 9, 52]. As a naïve Bayesian method, baySeq framework has been proposed in differential expression analysis between different groups with replicates [52]. ShrinkBayes has been developed to identify differential expression analysis at static data on the basis of zero-inflated Poisson Gamma model and Integrated Nested Laplace Approximation to estimate shrunken parameters [53]. And another Bayesian technique in ChipSeq, BayesPeak has been developed to detect significantly enriched regions in transcription factor binding sites and histone modification datasets. It is on the basis of Bayesian hidden Markov model strategy and MCMC simulations with Poisson Gamma distribution to take into account over-dispersion in the abundance of read counts in different regions by comparing to the existing methods of peak callers in ChipSeq data, MACS, PeakSeq, and ChipSeq Peak Finder [54]. Additionally, Bayesian approaches to define differential expression of alternative splicing in RNA-seq have been proposed by MISO (as known as mixture of isoforms) model and MATS (as known as multivariate analysis of transcript splicing) [9, 18, 55]. MISO method is to estimate differential expression of alternatively spliced exons and isoforms based on Bayes factor to quantify the odds of differential regulation of given isoform for the ratio of inclusion and exclusion levels. Similarly, MATS (and rMATS) is implemented to test differential alternative splicing patterns by estimating exon inclusion levels between two samples without (and with replicates), respectively. In addition, microRNAs target prediction methods that play a key regulatory role in gene regulation on the variety of biological processes in human diseases, especially, cancer development, have been proposed by Bayesian methods [56, 57]. And also, Bayesian network methods have been proposed to predict the important functions of long non-coding RNAs as well as coding genes in RNA-seq [58, 59]. Thus, RNA-seq has become the alternative in transcriptome studies with advantageous features over arrays, dynamic range of expression signals, higher reproducibility and quality on samples and upgraded annotation without any known priori [3, 16, 19, 20, 24, 26–28, 32].

With the strength on improvements on technology and continuously declining cost to sequencing, RNA-seq enables to facilitate to perform more dense experimental designs such as time course data with abundant resources of dynamic gene regulation. Furthermore, transcriptome data and other types of meta-framed data across different platforms will be more popularly investigated in years to come. Indeed, next generation sequencing technologies have been steadily improved with higher throughput, longer reads, deeper sequencings, larger samples size of replicates, and less biases on data. Such advances allow investigators to conduct more complex experimental studies, various types of time course experiments, such as single series of longitudinally measured time course and transient dynamic patterns in

developmental stages; multi-series of factorial time course with multiple external conditions at each time point; cell-cyclic periodical data with (or without) external stimuli [7, 8, 22, 23, 29, 30, 60].

To date, despite the substantial potential and significance to explore temporal dynamics at genes and other genomic features in various human disease progressive models and therapeutic effects, the lack of analytical methodologies in order to precisely characterize temporal dynamics has been an important challenging issue to better understand biological mechanisms relevant with both time specific and responsively altered changes by given external stimulus.

In this chapter, we propose Bayesian approaches to better infer differential expression in temporal and spatial dynamic regulation that can be widely adopted in the community of biomedical research such as pediatric disease progressive models, age-related neurodegenerative diseases, other types of longitudinal and multi-series of time course data.

## 2. Methods

### 2.1. Data types of time course experiments

Dynamic gene regulation within a time window in transcriptome data is generally subcategorized into (1) within-subject longitudinally repeatedly measured stimuli-response data in a single series of time course experiment, (2) between-subject factorial multi-series of time course data with different conditions at each time point, and (3) periodical data in cell-cycle or circadian rhythmic patterns with or without external conditions. For the first type of temporal dynamics, we initially proposed Bayesian Poisson Gamma (negative binomial model) strategy to identify temporally differentially expressed genes in the previous study [23]. In this model, each gene is statistically tested whether it is equal or differential expression by auto-regressive (AR) model. The detailed description for notations is given in the following,

**Model I:** supposedly, a gene expression profile across a series of multiple time points,  $y_{grt}$  is independently distributed as Poisson Gamma (negative binomial model to account for variability of biological replicates within a group),  $g = 1, 2, \dots, G$  (gene or other genomic feature),  $t = 1, 2, \dots, T$  (time point), and  $r = 1, 2, \dots, R$  (biological replicate at each time point),

$$y_{grt} \sim \text{POI}(\mu_{grt}),$$

$$\log(\mu_{grt}) = w_{grt} + \beta_{g'}$$

$$w_{gr1} \sim N\left(0, \frac{\sigma^2}{1 - \varphi_{gr1}^2}\right),$$

$$w_{grt} \mid w_{gr,1,2,\dots,t-1} \sim N(\varphi_{gr1} * w_{grt-1}, \sigma^2), t > 2$$

In this proposed model,  $\beta_g$  is assumed to have non-informative priors and time series random effects model for sequentially measured single series of time course data is assumed. To update the defined auto-regressive model and estimate the posterior probabilities of given parameter sets, we employ Markov Chain Monte Carlo simulations,  $N = 10,000$  iterations and 8000 burn-ins. To define temporal dynamics whether or not a given gene is temporally differentially expressed, we further examine the most interesting parameter of auto-coefficient representing time series sequential random effects in this proposed model. To classify between equal and differential expression, we implemented to compute Bayesian credible interval estimates. This proposed model is implemented by OpenBUGS (WinBUGS) in R (submitted paper). Basically, it allows us to include major factor of time and variability of replicates at each time point and this simple linear auto-regressive (AR) model is straightforwardly extended to identify temporally differentially expressed other genomic features, for instance, quantified abundance of transcripts. Despite the much better improved quality of samples in RNA-seq data when compared to the beginning of technology, the preprocessing and normalization procedures are still required to precisely infer temporally differentially expressed genes and isoforms and to reduce the misleading results in subsequent analyses. And some samples are discarded in the preprocessing step such as due to sample preparation in experiments and the rest of corresponding samples for the given missing sample should be also deleted as it is the method for longitudinal data with repeated measurements across the series of time points. More importantly, RNA-seq data is highly skewed expression data toward to zero and low expression levels than high expression level. Collectively, in the format of longitudinally measured experimental designs in temporal dynamics of RNA-seq data, as further improved strategy, we are currently developing a zero-inflated Poisson Gamma model with missing observations in longitudinal data to improve the capability of detection of temporal changes in highly skewed count data [61, 62]. The detailed descriptions and notations are given in the following, we consider the conditional distribution of  $y_{grt} \mid E_{grt}$

$$y_{grt} \mid E_{grt} \sim \text{POI}(\mu_{grt}), \text{ if } E_{grt} = 1$$

and

$$P(y_{grt} = 0 \mid E_{grt} = 0) = 1 \text{ if } E_{grt} = 0$$

where  $E_{grt}$  is the binary indicator of whether a gene expression profile across a series of multiple time points is present for a gene  $g$ , time  $t$  and replication  $r$ . We also assume that conditioned on  $p_{grt}$ , the  $E_{grt}$ 's are independent Bernoulli random variables with  $P(E_{grt} = 1) = p_{grt}$  for  $g = 1, 2, \dots, G$ ,  $t = 1, 2, \dots, T$ , and  $r = 1, 2, \dots, R$ . Here, given  $E_{grt} = 1$ , assume that the  $y_{grt}$ 's are conditionally independent.

Compared to arrays, the major strength of RNA-seq transcriptome data enables to quantify and identify spliced isoforms as well as individual exon-level expression which had not been previously done due to low resolution [8–11, 13, 14, 25, 28, 30, 33, 35, 60, 63–68]. In addition to gene level analyses, it is well established that alternative splicing is a prevalent mechanism on the variety of organisms. It involves multiple selective schemes of splice sites to construct diverse functional pathways and protein structures in gene regulation. A single mRNA may



code for different forms of a protein (isoform) as a result of alternative splicing that increases the complexity of mammalian transcriptome.

In the previous literatures, deep sequencing based transcriptome data predicts that more than ~95% of human genes typically contain multi-exons and transcript-variants to undergo alternative spliced events [8, 9, 25, 28, 33, 35, 63–65, 67]. The aberrant alternative splicing events occurring in post-transcriptional and translational procedures are highly associated with different tissues- or developmental stages- or environmental condition-specific manner. It has been investigated that malformation and dysfunctional mechanisms by the majority of abnormal alternative splicings in human brain diseases. Aberrant patterns of splicings in neurodegenerative Alzheimer's patients and other types of pediatric cancer progression could be significant contributors to targeted therapies on disease progressive models and developmental evolutionary processes in transcriptional activity in temporal dynamics [28, 63–66]. Despite the importance of alternative splicing events in recent technology, characterization of dynamic processes has been merely limited to gene level and static data approaches.

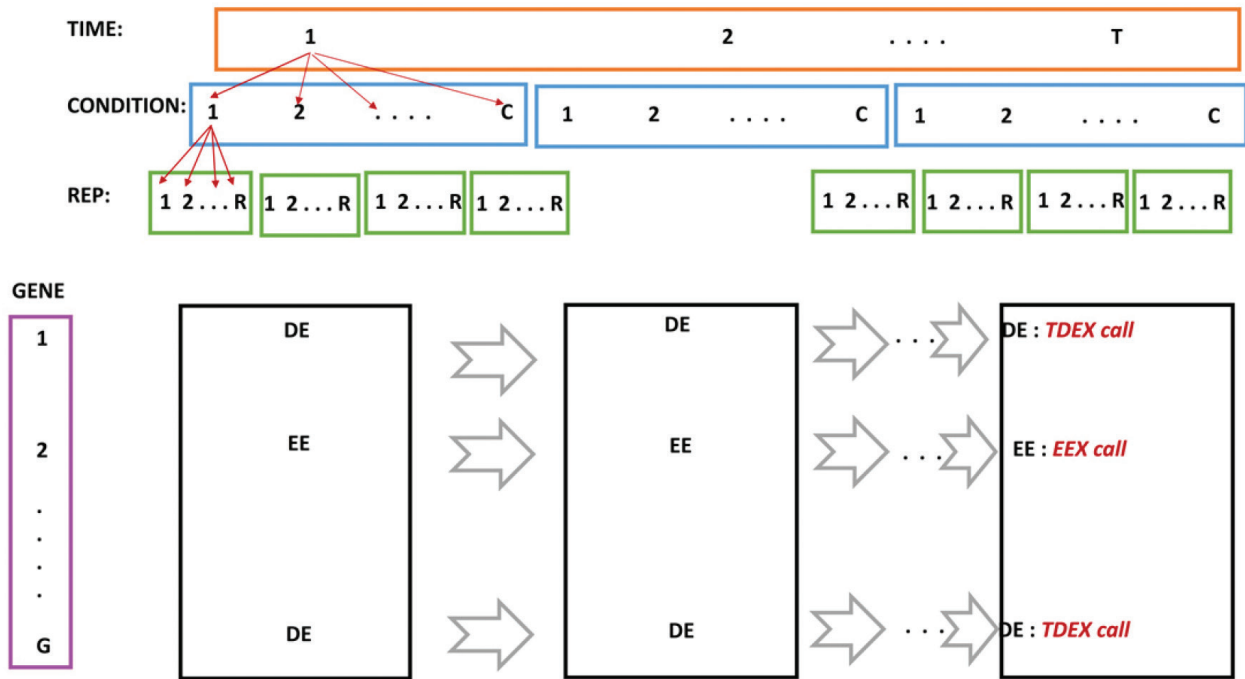
In the methodological point of view, for quantification and identification of isoforms, a couple of bioinformatics tools including IQSeq, rSeq, MapSplice have been recently developed [15, 17, 31], and identification of differentially expressed isoforms at static data types have been pursued by MATS (focused on a specified experimental design on a sample versus another single sample comparison at a fixed time point) [9], DEXSeq (flexibly to allow various types of experimental and biological conditions in generalized linear model from the basis of multiple comparisons at exon levels) [5], and cufflinks and cuffdiff (as known as one of the most popular versatile tools for quantification and identification of differential expression at isoform levels, but restricted to simple pairwise comparison with replicates) [10, 11, 30, 68].

To our best knowledge, none of current static and dynamic methods can identify temporally differentially expression at alternative splicing by explicitly accounting for data-driven nature of various time course experimental settings.

For the first type of longitudinal time course experiments, quantified expression levels at isoforms and other types of genomic features can be directly applied for our proposed dynamic AR model.

For the second type of between-subject factorial multi-series of time course data, another our previous study [8] proposed a hierarchical Bayesian modeling approach to define differential expression analysis at isoforms when having multiple conditions at each time point, such as different tissues, drug treatments, stress, and trauma in temporal dynamics (see **Figure 1**).

**Model II:** the detailed description for notations in the proposed model is given in the following, for the sake of consistency in notations of the proposed models, we pursue Bayesian hierarchical Dirichlet Bayesian mixture model of temporal dynamics in multi-series of time course data by making use of the identical notations on the transcriptomic expression levels as shown in the Model I. Suppose that a gene (or other genomic feature) expression profile across a series of multiple time points is tested and each time point contains different external conditions, such as different types of tissues, cell lines, drug treatments, trauma, stress, and et cetera, where we denote a condition factor,  $c = 1, 2, \dots, C$  at each time point.



**Figure 1.** It depicts the latent variable to be estimated in hierarchical Dirichlet Bayesian mixture model.

For a particular time point at  $T = t$ , and condition at  $C = c$ , the given expression level is given by  $y_{grtc}$   $g = 1, 2, \dots, G$  (gene),  $r = 1, 2, \dots, R$  (replicate at each time point and a condition),  $t = 1, 2, \dots, T$  (time point), and  $c = 1, 2, \dots, C$  (external condition at each time point). Let  $\gamma_g$  be a latent variable to be estimated in a hierarchical mixture model whether equal expression (EEX) or temporally differentially expression (TDEX) is at a given gene of interest. At a given time point  $T = t$ ,  $\tau_g$  denotes the  $g$ th gene effect,  $\beta_c$  denotes the effect of the  $c$ th condition. The latent variable  $\gamma_g$  represents an indicator variable such that the given  $g$ th gene is temporally differentially expressed or not, that is,  $\gamma_g = \begin{cases} 1, & \text{when TDEX} \\ 0, & \text{when EEX} \end{cases}$ .

$$y_{grtc} \mid \gamma_g = d \sim \text{POI}(\mu_{grtc}),$$

$$\log(\mu_{grtc}) = \tau_g + \beta_c + F(\beta) + w_{grtc},$$

$$\beta_c \mid l_g = l \sim F(\beta_l),$$

$l$  represents distinct patterns between conditions across time points,

Namely,  $l_g = l$  iff  $g$ th gene belongs to cluster  $l$ ,

Let  $L$  denote the number of clusters on expression data,

$n_g = |\{g; l_g = l\}|$  denote the size of the  $l$ th cluster,

$$F(\beta) \sim DP(F_0(F(\beta)); \eta_0),$$

$$w_{grtc} \mid (r_g = d, l_g = l) \sim DP(G_d^*; M),$$

where  $DP(H; \alpha)$  stands for the Dirichlet process having its baseline distribution  $H(\cdot)$  and mass parameter  $\alpha$ . And,  $F_0$  represents the mixture format of  $\beta$  with  $p(\beta) = N_{G-1}(0, \psi_0)$  such that  $G_0^* \sim N(0, \sigma^2)$  for equal expression (EEX) and  $G_1^* \sim \frac{1}{2}N(-\delta, \sigma^2) + \frac{1}{2}N(\delta, \sigma^2)$  for differential expression (DEX) and  $\eta_0, \psi_0$  and  $M$  are fixed hyper-parameters based on prior information [6, 69]. Therefore, the posterior probability is given by,  $P(\gamma_g = d \mid l_g = l, y_{g11}, \dots, y_{g, R, t=T, C=C})$  for the temporally differentially expressed gene.

Based on our proposed Bayesian approach for multi-series of factorial time course data, we are currently implementing OpenBUGS(WinBUGS) in R to perform differential expression analysis. In order to validate our proposed model, we need to compare to other maSigPro for RNA-seq data and Gaussian Process modeling approach in terms of temporally differentially expressed genes in the multiple datasets after transformation of stabilizing variance on counts data.

### 3. Closing remarks in future directions

In earlier sections, we have discussed Bayesian techniques to address different types of experimental (clinical) settings in temporal dynamics by focusing on Poisson Gamma autoregressive model for longitudinally measured single-series of time course data and hierarchical Dirichlet Bayesian mixture model framework for multi-series of factorial time course data, respectively. Thus, as the continuous efforts to modeling approaches, we propose differential expression analytical frameworks that precisely characterize temporal dynamics for each type of stimuli-response data in this chapter.

The novel features in the proposed hierarchical Dirichlet Bayesian mixture model enables to identify significant temporal changes of expression levels between at least two external conditional factors over a series of time points based on Bayesian strategy grouping clusters by the patterns of differential or equal expression [6, 69]. The identified temporal changes are determined as the putative biomarkers that could be relevantly linked with various dynamic genetic mechanisms of molecular and physiological processes. Additionally, our proposed method enables to allow more than two genetic and environmental factors within a time point and to address how the intra-factor of multiple conditions and time factor affect altered expression patterns as significant contributors independently and interactively. Furthermore, this proposed model is straightforwardly extended to detect temporal changes at other genomic levels such as transcripts and exon levels.



As the extension of this study, we are currently developing how to measure the relationship of a parent gene-to-multiple child isoforms in temporal dynamic patterns. For the task of this procedure, we carry out directional comparison, gene-to-isoform in differential expression based on similarity and discrepancy on magnitude and pattern of expression. This proposed model enables to define connectivity visualization of splicing maps on the variety of structural formations by switchable exon usage. Moreover, we are currently developing differential expression method for cell-cyclic periodical data with or without external conditions in stimuli-response data [70]. Thus, this proposed study is timely crucial to define temporal dynamics at alternative splicing diversity related with disease progression by discovering which splicing events are condition and time specifically observed and how eventually their spliced abnormal patterns and splicing maps are associated with biological functions. And it is essential to develop strategies to correct aberrant splicing as well as gene approaches in temporal and spatial dynamics on the variety of disease progression and evolutionary comparative studies between human diseases and other closely related species.

## **Conflict of interest**

The authors have no conflicts of interest to disclose.

## **Contributors' statements**

SO wrote manuscript and SO and SS conceived this study.

For discloser of any prior publications or submission with any overlapping information including studies and patients, there are no prior publications or submissions with any overlapping information including studies and patients.

The manuscript has not been and will not be submitted to any other journal while it is under consideration by this book chapter in Bayesian inference.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

## **Funding source**

This study is supported by an internal grant from Jeju National University to Dr. Sunghee Oh.

## **Financial disclosure**

The authors have no financial relationships relevant to this article to disclose.

## Author details

Sunghee Oh<sup>1\*</sup> and Seongho Song<sup>2</sup>

\*Address all correspondence to: [sshshshoh1105@gmail.com](mailto:sshshshoh1105@gmail.com)

1 Department of Computer Science and Statistics, Jeju National University, Jeju City, South Korea

2 Department of Mathematical Sciences, University of Cincinnati, Cincinnati, USA

## References

- [1] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*. 2010;**11**:R106. DOI: 10.1186/gb-2010-11-10-r106
- [2] Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010;**11**:94. DOI: 10.1186/1471-2105-11-94
- [3] Ritchie ME et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015;**43**:e47. DOI: 10.1093/nar/gkv007
- [4] Robinson MD, DJ MC, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;**26**:139-140. DOI: 10.1093/bioinformatics/btp616
- [5] Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Research*. 2012;**22**:2008-2017. DOI: 10.1101/gr.133744.111
- [6] Guindani M, Sepulveda N, Paulino CD, Muller P. A Bayesian semi-parametric approach for the differential analysis of sequence counts data. *Journal of the Royal Statistical Society. Series C, Applied Statistics*. 2014;**63**:385-404. DOI: 10.1111/rssc.12041
- [7] Heinonen M et al. Detecting time periods of differential gene expression using Gaussian processes: An application to endothelial cells exposed to radiotherapy dose fraction. *Bioinformatics*. 2015;**31**:728-735. DOI: 10.1093/bioinformatics/btu699
- [8] Oh S, Song S. Differential gene expression (DEX) and alternative splicing events (ASE) for temporal dynamic processes using HMMs and hierarchical Bayesian modeling approaches. *Methods in Molecular Biology*. 2017;**1552**:165-176. DOI: 10.1007/978-1-4939-6753-7\_12
- [9] Shen S et al. MATS: A Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Research*. 2012;**40**:e61. DOI: 10.1093/nar/gkr1291
- [10] Trapnell C et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*. 2013;**31**:46-53. DOI: 10.1038/nbt.2450

- [11] Trapnell C et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2012;**7**:562-578. DOI: 10.1038/nprot.2012.016
- [12] Anders S, Pyl PT, Huber W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;**31**:166-169. DOI: 10.1093/bioinformatics/btu638
- [13] Ayoub AE et al. Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;**108**:14950-14955. DOI: 10.1073/pnas.1112213108
- [14] Dasgupta N et al. Gaucher disease: Transcriptome analyses using microarray or mRNA sequencing in a Gba1 mutant mouse model treated with velaglucerase alfa or imiglucerase. *PLoS One*. 2013;**8**:e74912. DOI: 10.1371/journal.pone.0074912
- [15] Du J et al. IQSeq: Integrated isoform quantification analysis based on next-generation sequencing. *PLoS One*. 2012;**7**:e29175. DOI: 10.1371/journal.pone.0029175
- [16] Hansen KD, Wu Z, Irizarry RA, Leek JT. Sequencing technology does not eliminate biological variability. *Nature Biotechnology*. 2011;**29**:572-573. DOI: 10.1038/nbt.1910
- [17] Jiang H, Wong WH. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics*. 2009;**25**:1026-1032. DOI: 10.1093/bioinformatics/btp113
- [18] Katz Y, Wang ET, Airolidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*. 2010;**7**:1009-1015. DOI: 10.1038/nmeth.1528
- [19] Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 2008;**18**:1509-1517. DOI: 10.1101/gr.079558.108
- [20] Metzker ML. Sequencing technologies—The next generation. *Nature Reviews. Genetics*. 2010;**11**:31-46. DOI: 10.1038/nrg2626
- [21] Mills JD et al. RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neuroscience Letters*. 2013;**536**:90-95. DOI: 10.1016/j.neulet.2012.12.042
- [22] Nueda MJ, Tarazona S, Conesa A. Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics*. 2014;**30**:2598-2602. DOI: 10.1093/bioinformatics/btu333
- [23] Oh S, Song S, Grabowski G, Zhao H, Noonan JP. Time series expression analyses using RNA-seq: A statistical approach. *BioMed Research International*. 2013;**2013**:203681. DOI: 10.1155/2013/203681
- [24] Reis-Filho JS. Next-generation sequencing. *Breast Cancer Research*. 2009;**11**(Suppl 3):S12. DOI: 10.1186/bcr2431

- [25] Richard H et al. Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Research*. 2010;**38**:e112. DOI: 10.1093/nar/gkq041
- [26] Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*. 2014;**32**:896-902. DOI: 10.1038/nbt.2931
- [27] Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*. 2011;**12**:R22. DOI: 10.1186/gb-2011-12-3-r22
- [28] Sultan M et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*. 2008;**321**:956-960. DOI: 10.1126/science.1160342
- [29] Sun X et al. Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model. *BMC Bioinformatics*. 2016;**17**:324. DOI: 10.1186/s12859-016-1180-9
- [30] Trapnell C et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010;**28**:511-515. DOI: 10.1038/nbt.1621
- [31] Wang K et al. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*. 2010;**38**:e178. DOI: 10.1093/nar/gkq622
- [32] Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*. 2010;**11**:R14. DOI: 10.1186/gb-2010-11-2-r14
- [33] Zhao K, Lu ZX, Park JW, Zhou Q, Xing Y. GLiMMPS: Robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biology*. 2013;**14**:R74. DOI: 10.1186/gb-2013-14-7-r74
- [34] Zhou YH, Xia K, Wright FA. A powerful and flexible approach to the analysis of RNA sequence count data. *Bioinformatics*. 2011;**27**:2672-2678. DOI: 10.1093/bioinformatics/btr449
- [35] Ezkurdia I et al. Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Molecular Biology and Evolution*. 2012;**29**:2265-2283. DOI: 10.1093/molbev/mss100
- [36] Margolin AA et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*. 2006;**7**(Suppl 1):S7. DOI: 10.1186/1471-2105-7-S1-S7
- [37] Ronen M, Rosenberg R, Shraiman BI, Alon U. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;**99**:10555-10560. DOI: 10.1073/pnas.152046799
- [38] Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature Protocols*. 2015;**10**:823-844. DOI: 10.1038/nprot.2015.052

- [39] Zoppoli P, Morgarella S, Ceccarelli M. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics*. 2010;**11**:154. DOI: 10.1186/1471-2105-11-154
- [40] Chen R et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*. 2012;**148**:1293-1307. DOI: 10.1016/j.cell.2012.02.009
- [41] Loven J et al. Revisiting global gene expression analysis. *Cell*. 2012;**151**:476-482. DOI: 10.1016/j.cell.2012.10.012
- [42] Shah SP et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;**486**:395-399. DOI: 10.1038/nature10933
- [43] Martin CL et al. Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*. 2007;**144B**:869-876. DOI: 10.1002/ajmg.b.30530
- [44] Guirguis A et al. Use of gene expression profiles to stage concurrent endometrioid tumors of the endometrium and ovary. *Gynecologic Oncology*. 2008;**108**:370-376. DOI: 10.1016/j.ygyno.2007.10.008
- [45] Kalyana-Sundaram S et al. Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia*. 2012;**14**:702-708
- [46] McElwee JL et al. Identification of PADI2 as a potential breast cancer biomarker and therapeutic target. *BMC Cancer*. 2012;**12**:500. DOI: 10.1186/1471-2407-12-500
- [47] Oh S, Tseng GC, Sibille E. Reciprocal phylogenetic conservation of molecular aging in mouse and human brain. *Neurobiology of Aging*. 2011;**32**:1331-1335. DOI: 10.1016/j.neurobiolaging.2009.08.004
- [48] Sibille E et al. A molecular signature of depression in the amygdala. *The American Journal of Psychiatry*. 2009;**166**:1011-1024. DOI: 10.1176/appi.ajp.2009.08121760
- [49] Chiu IM et al. A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Reports*. 2013;**4**:385-401. DOI: 10.1016/j.celrep.2013.06.018
- [50] Xu YH, Sun Y, Barnes S, Grabowski GA. Comparative therapeutic effects of velaglucerase alfa and imiglucerase in a Gaucher disease mouse model. *PLoS One*. 2010;**5**:e10750. DOI: 10.1371/journal.pone.0010750
- [51] Cotney J et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Research*. 2012;**22**:1069-1080. DOI: 10.1101/gr.129817.111
- [52] Hardcastle TJ, Kelly KA. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*. 2010;**11**:422. DOI: 10.1186/1471-2105-11-422



- [53] van de Wiel MA, Neerincx M, Buffart TE, Sie D, Verheul HM. ShrinkBayes: A versatile R-package for analysis of count-based sequencing data in complex study designs. *BMC Bioinformatics*. 2014;**15**:116. DOI: 10.1186/1471-2105-15-116
- [54] Spyrou C, Stark R, Lynch AG, Tavaré S. BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*. 2009;**10**:299. DOI: 10.1186/1471-2105-10-299
- [55] Shen S et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;**111**:E5593-E5601. DOI: 10.1073/pnas.1419161111
- [56] Liu H et al. A Bayesian approach for identifying miRNA targets by combining sequence prediction and gene expression profiling. *BMC Genomics*. 2010;**11**(Suppl 3):S12. DOI: 10.1186/1471-2164-11-S3-S12
- [57] Wang Z, Xu W, Zhu H, Liu Y. A bayesian framework to improve microRNA target prediction by incorporating external information. *Cancer Informatics*. 2014;**13**:19-25. DOI: 10.4137/CIN.S16348
- [58] Xiao Y et al. Predicting the functions of long noncoding RNAs using RNA-seq based on Bayesian network. *BioMed Research International*. 2015;**2015**:839590. DOI: 10.1155/2015/839590
- [59] van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in Bioinformatics*. 2017. DOI: 10.1093/bib/bbw139
- [60] Oh S, Song S, Dasgupta N, Grabowski G. The analytical landscape of static and temporal dynamics in transcriptome data. *Frontiers in Genetics*. 2014;**5**:35. DOI: 10.3389/fgene.2014.00035
- [61] Neelon BH, O'Malley AJ, Normand SL. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*. 2010;**10**:421-439. DOI: 10.1177/1471082X0901000404
- [62] Wang X, Chen MH, Kuo RC, Dey DK. Bayesian spatial-temporal modeling of ecological zero-inflated count data. *Statistica Sinica*. 2015;**25**:189-204. DOI: 10.5705/ss.2013.212w
- [63] Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. *Nature Biotechnology*. 2004;**22**:535-546. DOI: 10.1038/nbt964
- [64] Mills JD, Janitz M. Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiology of Aging*. 2012;**33**(1012):e1011-e1024. DOI: 10.1016/j.neurobiolaging.2011.10.030
- [65] Sanford JR, Ellis JD, Cazalla D, Caceres JF. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;**102**:15042-15047. DOI: 10.1073/pnas.0507827102



- [66] Stower H. Splicing: Waiting to be spliced. *Nature Reviews. Genetics*. 2012;**13**:599. DOI: 10.1038/nrg3310
- [67] Wang ET et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;**456**:470-476. DOI: 10.1038/nature07509
- [68] Kim D et al. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013;**14**:R36. DOI: 10.1186/gb-2013-14-4-r36
- [69] Medvedovic M, Sivaganesan S. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*. 2002;**18**:1194-1206
- [70] Spellman PT et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*. 1998;**9**:3273-3297

IntechOpen